Evolution and Human Behavior xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Explaining the evolution of parochial punishment in humans

Miguel dos Santos*, Daria Knoch

Department of Social Neuroscience and Social Psychology, Institute of Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland

ARTICLE INFO

Keywords:
Parochialism
In-group bias
Cooperation
Intergroup conflict
Group competition
Human behaviour

ABSTRACT

Humans usually favour members of their own group, ethnicity or culture (parochial cooperation), and punish out-group wrongdoers more harshly (parochial punishment). The evolution of parochial cooperation is mainly explained by intergroup conflict, as restricting cooperation to in-groups can provide a relative advantage during conflict. However, explanations for the evolution of parochial punishment are still lacking. It is unclear whether conflict can also explain parochial punishment, because conflict is expected to lead to full hostility towards out-groups, irrespective of their behaviour. Here, we use an agent-based simulation to explore which conditions favour the evolution of parochial third-party punishment. We show that when groups interact and then engage in conflict with each other, third-party punishment is not parochial but spiteful, and directed towards all out-groups. A parochial bias in punishment decisions evolves (i) without conflict, when groups compete against nature and enforcing cooperation requires many punitive actions, and (ii) with conflict, when groups come into conflict with a group other than one they previously interacted with. Our findings suggest that intergroup conflict does not always lead to parochial punishment, and that stable collaborative relations between groups is a key factor promoting third-party parochial punishment. Our findings also provide novel predictions on how punishment and intergroup conflict influence in-group bias in human societies.

1. Introduction

Humans usually have a tendency to favour members of their own group, ethnicity or culture. This kind of in-group bias is usually referred to as parochialism (or ethnocentrism, in-group favouritism), and it is often associated with cautiousness about, or even hostility towards outsiders (Bernhard, Fehr, & Fischbacher, 2006 Bernhard, Fischbacher, & Fehr, 2006 Brewer, 1999 Hewstone, Rubin, & Willis, 2002). Parochialism has also been well documented in many laboratory experiments, where participants are divided into subgroups based on either real life or artificial groupings (Balliet, Wu, & De Dreu, 2014 Efferson, Lalive, & Fehr, 2008 Everett, Faber, & Crockett, 2015 Götte, Huffman, & Meier, 2006 Ockenfels & Werner, 2014 Romano, Balliet, Yamagishi, & Liu, 2017 Tajfel, Billig, Bundy, & Flament, 1971). In these experiments, people typically donate more to in-group partners, and entrust them with more resources.

A second important facet of human parochialism is that people also use parochial punishment: they invest resources to punish uncooperative partners (cheats), and do so more severely if cheats come from a different group than their own (Baumgartner, Götte, Gügler, & Fehr, 2012 Baumgartner, Schiller, Rieskamp, Gianotti, & Knoch, 2014 Bernhard, Fehr, & Fischbacher, 2006; Bernhard, Fischbacher, & Fehr, 2006; Delton & Krasnow, 2017 Guo, Ding, & Wu, 2020 Martin, Young, & McAuliffe, 2020 McAuliffe & Dunham, 2016 Mussweiler & Ockenfels, 2013 Schiller, Baumgartner, & Knoch, 2014 Yudkin, Rothmund, Twardawski, Thalla, & Van Bavel, 2016). Parochial punishment develops early in development, as it can be found in children as young as six years old (Jordan, McAuliffe, & Warneken, 2014). It has also been observed in experiments with real groupings such as army platoons (Götte et al., 2006 Götte, Huffman, Meier, & Sutter, 2012) and students associations (Liu et al., 2018), as well as between tribe members in Papua New Guinea (Bernhard, Fehr, & Fischbacher, 2006; Bernhard, Fischbacher, & Fehr, 2006). The ubiquitous nature of parochial cooperation and punishment in humans suggests that these behaviours evolved as part of our ancestors' psychology when early humans lived in small-scale societies (De Dreu, Balliet, & Halevy, 2014 Delton & Krasnow, 2017 Krasnow, Delton, Tooby, & Cosmides, 2013 McAuliffe & Dunham, 2016 Yudkin et al., 2016). Hence, explaining their evolution is of major importance for fully understanding human behaviour in groups, and in turn, for a potential reduction in the impact of intergroup biases in human societies.

To explain the evolution of parochialism, researchers have mainly focused on hostile intergroup interactions (Böhm, Rusch, & Baron, 2018 Bowles, 2009 De Dreu et al., 2014, 2010 Masuda & Fu, 2015 McDonald,

* Corresponding author.
E-mail address: miguel.dossantos@psy.unibe.ch (M. dos Santos).

https://doi.org/10.1016/j.evolhumbehav.2020.10.002
Received 6 December 2019; Received in revised form 5 October 2020; Accepted 6 October 2020

Navarrete, & van Vugt, 2012 Puurtinen, Heap, & Mappes, 2015 Rusch, 2014). Intergroup conflict was probably frequent during our evolutionary history (Bowles, 2009 Choi & Bowles, 2007 Lahr et al., 2016). If groups compete with each other over territory or resources, then cooperating only with in-groups and weakening rival out-groups can provide a relative advantage during conflict. Many theoretical studies have indeed shown how intergroup conflict can promote parochial cooperation and its coevolution with between-group aggression (Choi & Bowles, 2007 García & van den Bergh, 2011 Konrad & Morath, 2012 Lehmann, 2011 Lehmann & Feldman, 2008). Punishment is usually not present in these studies, but they generally show that conflict between groups increases the selection pressure to help in-groups, because helpers can increase their own group's survival during war.

However, explanations for the evolution of parochial punishment are still lacking. It is unclear whether intergroup conflict can also explain parochial punishment, because conflict is expected to favour individuals who show hostility or aggression whenever they interact with out-groups. Hence, conflict is likely to favour spiteful, rather than parochial punishment, whereby both out-group cheats and cooperators are severely punished, which has been confirmed experimentally (Götte et al., 2012). In fact, theoretical studies on the evolution of parochialism typically limit intergroup interactions to hostile relations, preventing cooperation with out-groups from evolving (Gao, Wu, Nie, & Wang, 2015 Gao, Wu, & Wang, 2015 Konrad & Morath, 2012 Lehmann & Feldman, 2008). This is at odds with observations in real human societies, where positive interactions can occur between groups (Bowles, 2008 Pisor & Surbeck, 2019 Schaub, 2017), and with empirical findings showing that people (i) cooperate with out-groups, although to a lesser extent than with in-groups (Bernhard, Fehr, & Fischbacher, 2006; Bernhard, Fischbacher, & Fehr, 2006; Romano et al., 2017)), (ii) punish out-group cooperators much less frequently than out-group cheats (Baumgartner et al., 2012 Bernhard, Fehr, & Fischbacher, 2006; Bernhard, Fischbacher, & Fehr, 2006; Guo et al., 2020 Schiller et al., 2014), and (iii) do not always show intergroup hostility (Aaldering, Ten Velden, van Kleef, & De Dreu, 2018 De Dreu et al., 2014 Halevy, Weisel, & Bornstein, 2012 Koopmans & Rebers, 2009 Liu et al., 2018). Explaining how this behavioural pattern can evolve is necessary for fully understanding the evolution of human parochialism.

We ask which evolutionary mechanism can favour individuals who punish out-group cheats more severely than in-group cheats. To answer this question, we use an agent-based simulation. In our model, individuals interact in pairs in a donor game with many partners, which can be in- and out-group members. We focus on third-party punishment whereby individuals can observe a pair and potentially punish the donor, even though they were not directly affected by the donor's decision. Individuals can use third-party punishment with varying severity, and they can also change their behaviour in response to the punishment inflicted on either themselves or members of their own group. We model punishment as a continuous variable, and we allow individuals to condition their decisions based on their partner's group affiliation. This allows us to identify conditions under which the severity of punishment is greater for out-groups than in-groups. Importantly, we consider punishment to be parochial when it is targeted at in- and out-group cheats, and used to enforce cooperation. In contrast, we consider punishment to be spiteful when it is used to merely weaken out-groups, regardless of cooperation (Götte et al., 2012).

We explore four different scenarios. First, we consider a simple scenario in which there is no intergroup conflict, and individuals interact exclusively with in-groups. Here, groups face extinction due to environmental catastrophes, and their chance of survival is proportional to their group average payoff. Therefore, cooperation increases group survival. On the other hand, punishment decreases survival by decreasing group average payoff. However, punishment can also potentially deter future cheating, which could ultimately increase group survival. This scenario provides a baseline for when cooperation and punishment can evolve in our simulation. Second, we let individuals interact with members of another group, which provides scope for between-group cooperation and punishment. In this case, third-party punishers can potentially increase their own group survival by preventing out-groups from cheating with their own group members. This scenario enables us to test whether parochial third-party punishment can evolve even in the absence of intergroup conflict.

In the last two scenarios, we introduce physical conflict between groups, so that groups with a higher average payoff than their opponent are more likely to win and survive. Hence, conflict creates incentives to cooperate with in-groups and undermine out-groups. In the third scenario, conflict occurs between groups that previously interacted with each other. This allows us to test the previous verbal argument that intergroup conflict leads to parochial punishment (Choi & Bowles, 2007). Finally, we explore a fourth scenario in which groups are more likely to come into conflict with a group other than the one they previously interacted with. This condition allows us to narrow down which cases of intergroup cooperation and conflict can potentially lead to the evolution of parochial punishment.

2. Methods

2.1. Life cycle

We model a haploid population subdivided into $n_g = 100$ groups, each of size n (with $n_g \times n = N_T$). The life cycle is as follows: (i) adults accumulate payoffs during social interactions. (ii) Each group dies with some probability (either due to the environment or conflict with another group), depending on the group average payoff. (iii) Surviving individuals reproduce sexually, in proportion to their payoff acquired during the social interactions, with recombination and mutations. (iv) All offspring are randomly distributed into n_g groups (no population structure, global competition), and start the life cycle again (Fig. S1 in the supplementary information).

2.2. Social interactions

Individuals interact in pairs, and a random third-party from the local group can observe the interaction (Fig. 1a). There are three roles: the donor, the recipient and the third-party punisher. The donor can decide whether to give a benefit b to the recipient, at personal cost c (with $b > c$). After the donor's decision, a random third-party punisher who observed the interaction then invests some punishment effort P_i (continuous number between 0 and 1) to pay αP_i for the donor to lose βP_i (with punishment technology $\beta{:}\alpha$). Neither the donor, nor the recipient knows in advance the third-party punisher's identity. So, donors cannot condition their decision on the third-party punisher's past behaviour. Punishers can punish donors for giving and/or refusing to give (antisocial and justified/moralistic punishment, respectively). Interactions are repeated over R_{Tot} rounds, and in each round three individuals in each group are randomly chosen to form a subgroup. We use a large number of rounds ($R_{Tot} = 600$) to explore the area of parameter space where punishment can potentially promote cooperation.

We vary whether individuals interact with members of their own and/or a neighbour group (always the same neighbour group). The term ‘neighbour group’ is only used for presentation purposes, as there is no spatial structure in our model. We assume that individuals can sometimes ‘move’ to the neighbour group for an interaction either as the donor or recipient (Fig. 1b) and then come back to their own group. So, third-parties are always from the local group. Therefore, only three types of subgroup can occur, and we denote the subgroup combinations according to the punisher's point of view (Fig. 1b): (i) the punisher is from the same group as that of both the donor and recipient (In−In); (ii) the punisher is from the same group as that of the donor, whilst the recipient is an out-group (In-Out); and (iii) the punisher is from the

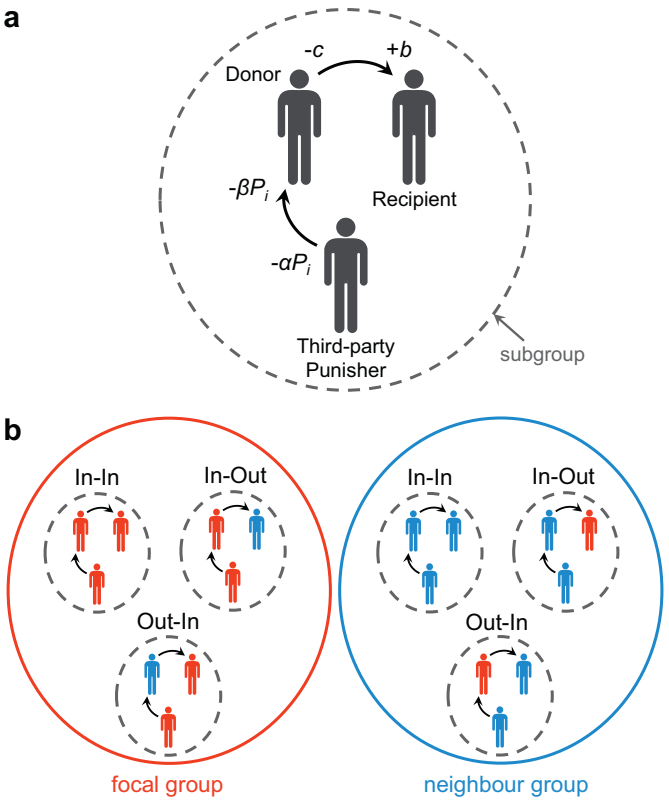

Fig. 1. Social interactions.
a, In each round, three randomly selected individuals form a subgroup and interact with each other. The donor can decide whether to pay a cost c for the recipient to receive a benefit b. The punisher observes this interaction and invests an effort P_i to punish the donor (for cooperating and/or defecting) and pays a cost αP_i. The donor pays a punishment fine βP_i (the punishment technology is thus β:α). **b**, Individuals from a focal group can interact with members of their own and a neighbour group (always the same neighbour group). We consider three types of subgroups: (i) the donor and recipient are both from the same group as the punisher (In−In), (ii) only the donor is from the same group as the punisher (In-Out), and (iii) only the recipient is from the same group as the punisher (Out−In).

same group as that of the recipient, whilst the donor is an out-group (Out−In). For simplicity, we did not include Out-Out subgroups in our simulation to reduce the already high number of traits individuals can carry. In this Out-Out situation, we expect full punishment of both out-group cheats and cooperators when groups engage in conflict with the neighbour group. If groups engage in conflict with a random group, we expect no selection (drift) for the punishment of out-group cheats, but parochial punishment might emerge as a result of selection for low punishment of in-group cheats, such as when in-group cooperation is difficult to enforce.

The number of rounds of each subgroup combination is given by $R_{In\text{-}In}$, $R_{In\text{-}Out}$ and $R_{Out\text{-}In}$ (with $R_{In\text{-}In} + R_{In\text{-}Out} + R_{Out\text{-}In} = R_{Tot}$). In order to avoid negative payoffs, we add a baseline payoff $c + \beta$ to all three individuals in each subgroup (i.e., donor, recipient and punisher). Therefore, on average individuals have a baseline payoff $f = 3R_{Tot}(c + \beta)/n$. This baseline payoff f is achieved, for example, when all group members are cheats who never punish.

2.3. Strategies

Individuals carry five social traits for each subgroup composition: (1) one binary trait codes for an individual's default cooperative behaviour when in the donor role (either C or D). This default behaviour will be expressed until the individual changes her behaviour after punishment (see below). (2–3) Two punishment traits code for an individual's effort invested into punishing donors who cooperate and cheat, respectively. These two punishment traits can vary on a continuous scale between 0 and 1 (inclusive). (4–5) Finally, two traits code for an individual's response to the punishment of cooperation and cheating, respectively. These two traits can vary continuously between 0 and 1 (inclusive).

The two traits coding for the response to punishment each represent an internal threshold which works as follows. If an individual's current behaviour when in the donor role is to cooperate (cheat), she will switch to cheating (cooperation) in all subsequent rounds as soon as a member of her own group (including herself) is punished after cooperating (cheating) by a punishment effort greater than her internal threshold. In other words, individuals will switch to the opposite behaviour after observing their current behaviour being punished sufficiently harshly. For example, consider an individual whose default (i.e., initial) behaviour is to cheat, and whose internal thresholds for justified and antisocial punishment are 0.3 and 0.8, respectively. This individual will cheat when in the donor role, but she will start to cooperate as soon as a cheat in her group (including herself) is punished with an effort > 0.3 (now, her current behaviour will be to cooperate). She will cooperate in the following rounds, but she will start cheating again as soon as a cooperator in her group (or herself) is punished with an effort > 0.8, and so on.

Individuals cannot switch their behaviour as a donor if they are themselves inflicting punishment. Because we assume that donors cannot know in advance the observing third-party's identity, we do not model a psychology in which individuals remember who punished or not. In contrast, we let donors the possibility to evolve a psychology that takes into account whether a particular decision (either C or D) was punished in their group. This reflects a situation in which individuals can learn via either direct observation or communication/gossip which behaviour has been punished to infer their own future probability of being punished. This assumption seems plausible because people often use what happens to others to infer what will happen to them (Krasnow, Delton, Cosmides, & Tooby, 2016). Furthermore, people tend to perceive members from a different group as similar to each other (out-group homogeneity effect; (Ostrom & Sedikides, 1992)), which is line with our assumption that individuals extrapolate the punishment behaviour of a single out-group to any out-group. However, this type of psychology can be conducive to cooperation (as punishing one cheat can potentially convert all other cheats in the group at the same time). Therefore, we later check the robustness of our results to this particular assumption, by assuming that only the punished agent can change her behaviour.

We assume that donors know their receiver's group affiliation, and whether they are interacting at their own group location or that of the

neighbour group (Fig. 1b). We also assume that punishers know the group affiliation of both the donor and receiver. Therefore, donors and punishers can condition their behaviour based on group membership. As a consequence, individuals carry the five social traits presented above, for each subgroup combination, resulting in a total of 15 traits. These traits can evolve independently due to recombination during reproduction and independent mutations.

2.4. Group survival

After the social interactions, groups die with some probability, in which case all individuals in the group die. Group survival depends on the group average payoff acquired during social interactions. Therefore, a focal individual can increase its own group's survival, and in turn, its own survival, by cooperating (for an interpretation based on 'multi-level' or 'group selection' see e.g., (Okasha, 2007; Wilson & Sober, 1994)). On the other hand, punishment decreases group average payoff, and hence, decreases group survival, unless punishment induces group members to cooperate in future interactions.

We vary whether there is intergroup conflict. In the absence of conflict, a focal group g survives environmental catastrophes with probability $s^{NC}(\overline{\pi}_g) = 1/\{1 + \exp[-h(\overline{\pi}_g - f)/(\pi_{max} - f)]\}$, where $\overline{\pi}_g = \frac{1}{n}\sum_i^n \pi_{i,g}$ is the average payoff in group g, and $\pi_{max} = f + R_{Tot}(b - c)/n$ is the maximum average payoff that a given group can achieve (full cooperation, no punishment). The parameter h determines the shape of the survival curve, with a fixed 0.5 chance of survival when the group's average payoff equals f (Fig. S2a in the supplementary information). The average payoff f is achieved, for example, when all group members are cheats who never punish. This group survival function ensures that the increase in survival as a function of the group average cooperation rate in the absence of punishment will be similar for any benefit b. This allows us to compare simulation runs with different benefits b. Yet, b will still matter for between-individual payoff differences. To avoid a population crash, if all groups die in a generation, two groups are randomly selected to survive (in proportion to their average payoff).

In the presence of intergroup conflict, each group comes into physical conflict with another one, and only the winning group in each pair survives. We vary whether the opponent group is the 'neighbour' group with which a focal group previously interacted socially, or another randomly selected group. The probability that group g wins a conflict against group k is proportional to the difference in average payoff between the two groups, and is given by $s^C(\overline{\pi}_g, \overline{\pi}_k) = 1/[1 + \exp(\overline{\pi}_k - \overline{\pi}_g)]$ (Fig. S2b in the supplementary information).

2.5. Reproduction

Individuals in the surviving groups reproduce sexually. They produce a total of $2N_T$ haploid gametes in proportion to their individual payoff. We add 0.001 to the payoff of all surviving individuals to ensure strictly positive payoffs. Recombination occurs as follows. All gametes are randomly paired, and fuse to form N_T juveniles. During fusion, each trait is transmitted from one of the two parents at random. Each trait mutates independently from the others with probability $\mu = 0.01$. If a mutation occurs in a continuous trait, a normally distributed error with mean 0 and standard deviation 0.1 is added to the current trait value. A mutation in the trait coding for the default cooperative behaviour switches the trait value from C to D or from D to C.

We start all our simulations with a population of cheats that never punish and never react to punishment. Table S1 in the supplementary information shows all the model parameters. Each simulation run lasted 10^6 generations. For each simulation run, we recorded the trait averages every 10 generations, and computed the long-term trait averages across the last 500′000 generations.

2.6. Data and code availability

The simulation was coded in MATLAB R2017b. The code and all the simulated data used in this paper are publicly available from the OSF data repository (DOI: 10.17605/OSF.IO/4XG5S).

3. Results

3.1. Within-group interactions and no conflict

We first consider a simple scenario where interactions take place only within groups, and where there is neither punishment nor intergroup conflict (i.e., no aggression). This allows us to determine the level of default cooperation that evolves in the absence of other forces. A single mutant who cooperates unconditionally when in the donor role can invade a population of cheats if the increase in group survival due to cooperating with others can sufficiently outweigh the cooperation cost (Equ. S7 in the supplementary information). However, cooperation does not go to fixation, as cooperators and cheats coexist under a large area of parameter space (Fig. S3a).

We next allow within-group third-party punishment to evolve. In this case, punishment allows cooperation to invade completely, provided the punishment ratio is large enough (Fig. S3b). Because cooperation is common and mutant cheats are rare, the effort in punishing cheats evolves mainly through drift, with a long-term average effort around 0.5 (Fig. S4e; Fig. S5). On the other hand, there is strong selection against antisocial punishment (Fig. S4d; Fig. S5), because enforcing cooperators to switch to defection decreases group survival. Accordingly, there is a high switching threshold for antisocial punishment, such that cooperators continue cooperating even if strongly punished (Fig. S4b).

Here, third-party punishers gain direct fitness benefits from inducing cheats to cooperate in future interactions. This increased cooperation in turn increases group survival against catastrophes. Therefore, the present result provides an additional example of how third-party punishment could have evolved in humans for self-serving purposes (Delton & Krasnow, 2017 Jordan & Rand, 2017 Krasnow et al., 2016, 2013 Roos, Gelfand, Nau, & Carr, 2013).

3.2. Between-group interactions and no conflict

We now let individuals of a given group interact with members of a neighbour group (always the same neighbour group). We find no parochial punishment in this case, as there is no difference in the effort invested in the punishment of in- and out-group cheats (Fig. 2b; Fig. S6). We also find that within-group cooperation now coevolves with between-group cooperation, but only when the donor is an out-group (Out−In; Fig. S7a-c). Because there is full cooperation in the In−In and Out-In subgroups, efforts in punishing cheats in both subgroups vary through drift, although they remain slightly higher than 0.5 (Fig. 2b and b; Fig. S6). These results therefore suggest that, with the particular psychology we assumed here, parochial third-party punishment does not evolve in peaceful times, and that conflict between groups might still be key in explaining its evolution.

The fact that we find no cooperation in In-Out subgroups, where only the recipient is an out-group, is not surprising (Fig. S7b). In our model, punishing an in-group for not giving to an out-group is actually detrimental to the punisher's group survival. In turn, cooperation with out-group recipients is not enforced. However, in Out-In subgroups cooperation is enforced only under sufficiently large punishment ratios (Fig. S7c). Additional simulations show that between-group cooperation in Out-In subgroups remains below 0.4 if punishment is not allowed to evolve (Fig. S8c). Hence, third-party punishment of out-group members paves the way for between-group cooperation in this scenario. Furthermore, we find that decreasing the proportion of between-group interactions leads to a between-group cooperation break down (Fig. S9).

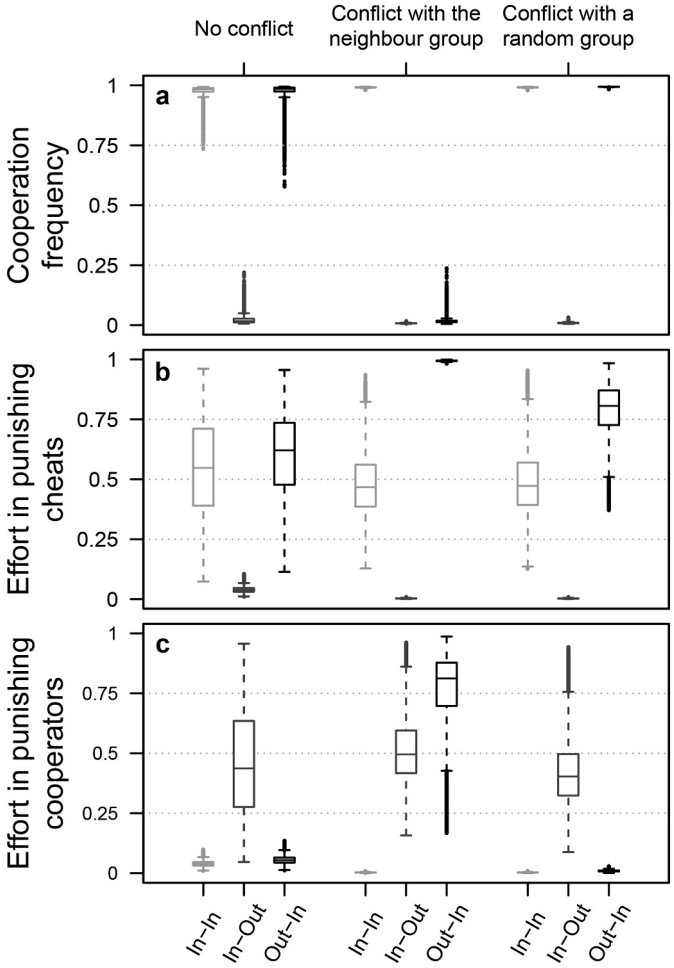

Fig. 2. Between-group collaboration is key for the evolution of parochial punishment.
Long-term range of the frequency of cooperation (**a**), effort in punishing cheats (**b**) and cooperators (**c**) for each subgroup composition in each ecological scenario. Data shown are from a parameter combination representative of most of the parameter range (see Fig. S6). In all panels, boxplots show the interquartile range and median over the last 500,000 generations of a simulation run. Parameters: $n = 20$, $n_g = 100$, $h = 4$, $R_{In\text{-}In} = 200$, $R_{Out\text{-}In} = 200$, $R_{In\text{-}Out} = 200$, $c = 1$, $b = 3$, $\alpha = 2$, $\beta = 6$, $\mu = 0.01$.

3.3. Intergroup conflict

We now assume that pairs of neighbouring (interacting) groups engage in conflict (physical aggression) after the social interactions, and that only individuals of the winning group survive. We find spiteful, rather than parochial punishment in this case. Even though individuals punish out-group cheats more harshly than they do in-groups (Fig. 2b; Fig. S6), out-group cooperators are also often punished, and there is no between-group cooperation (Fig. 2c; Fig. S7e and f). Hence, this type of punishment is used to harm all out-groups, rather than to enforce cooperation specifically. Furthermore, the switching threshold for justified punishment is maximal (Fig. S10), such that out-group cheats never switch to cooperation after being punished. The punishment of in-group cheats is even counter-selected when the benefits of cooperation are small and the punishment ratio is large (Fig. S6), suggesting that, with conflict, punishing rare in-group cheats is detrimental to group survival. As the punishment of in-group cooperators is also detrimental to group survival in the presence of conflict, there is also strong selection against it (Fig. 2c).

The average effort in punishing out-group cooperators remains high in this scenario when the cooperation benefit is small and the punishment ratio is high (Fig. S11, top right panels). So, there is still selection for punishing mutant out-group cooperators in this area of parameters space. We find similar results when decreasing the proportion of between-group interactions (Fig. S12), which means there is selection for punishing all out-group members even when between-group interactions are rare. These findings suggest that intergroup conflict promotes all-out hostility towards out-groups. This is in line with previous models, which found that within-group cooperation coevolves with between-group aggression (Choi & Bowles, 2007 Gao, Wu, Nie, & Wang, 2015; Gao, Wu, & Wang, 2015 Lehmann & Feldman, 2008). However, these models did not include parochial punishment, and thus could not explain its emergence.

We also examined the level of default cooperation that evolves in this scenario when punishment is prevented from evolving. We find that within-group cooperation fully invades even if punishment is not allowed to evolve (Fig. S8d-f). So, physical conflict between groups is sufficient to promote within-group cooperation. However, the mere presence of conflict with another group does not provide a satisfying explanation for parochial punishment. There must be some other factor favouring between-group cooperation, whilst mitigating the punishment of in-group cheats.

3.4. Collaborating and conflicting groups

We now examine what happens when conflict occurs with a random group, and hence, not necessarily with the neighbour group. We find that parochial third-party punishment evolves in this scenario. This is shown by three different results. First, individuals punish out-group cheats more harshly than they do in-groups (Fig. 2b; Fig. S6). Second, between-group cooperation evolves fully, even in the presence of intergroup conflict (Fig. S7i). Third, the punishment of out-group cooperators is strongly selected against (Fig. 2c; Fig. S11), unlike in the previous scenario where conflict occurred between interacting groups (Fig. S11).

The type of punishment that evolves here is parochial, as opposed to spiteful, as it is not used for aggression towards all out-groups. Instead, it is used to discipline out-group cheats. The punishment of rare out-group cheats remains high because it ensures that all potential cheats can be turned into cooperators, which is beneficial for a focal group's success against other groups. Furthermore, we find that under high punishment efficiencies, the effort invested in punishing in-group cheats is lower than that in the absence of conflict (Fig. S6). We find similar results when the proportion of between-group interactions is low (Fig. S13). This demonstrates that during conflict, the punishment of in-group cheats is moderate, because it would otherwise undermine group success.

3.5. Robustness

In the previous sections, we investigated situations in which conflict was either present or absent. Here, we check the robustness of our results by investigating situations in which each pair of groups comes into conflict with some probability. If a given pair of groups does not engage in conflict, both groups face extinction against the environment as in our first scenario. As before, we find that parochial punishment only evolves when conflict occurs with a random group. However, we find that parochial punishment becomes stronger as the probability of conflict increases (Fig. 3b; Fig. S14). Fig. 3 also shows that increasing the probability of conflict lowers the severity of the punishment of in-group cheats, suggesting that natural selection favours punishers who are more tolerant with their group members in times of conflict. Altogether, these results demonstrate that conflict creates selection pressure to punish out-group cheats severely in our simulation.

So far, we have assumed that not only the victims of punishment could adjust their behaviour after being punished, but also everyone else in the group. This particular psychology is conducive to

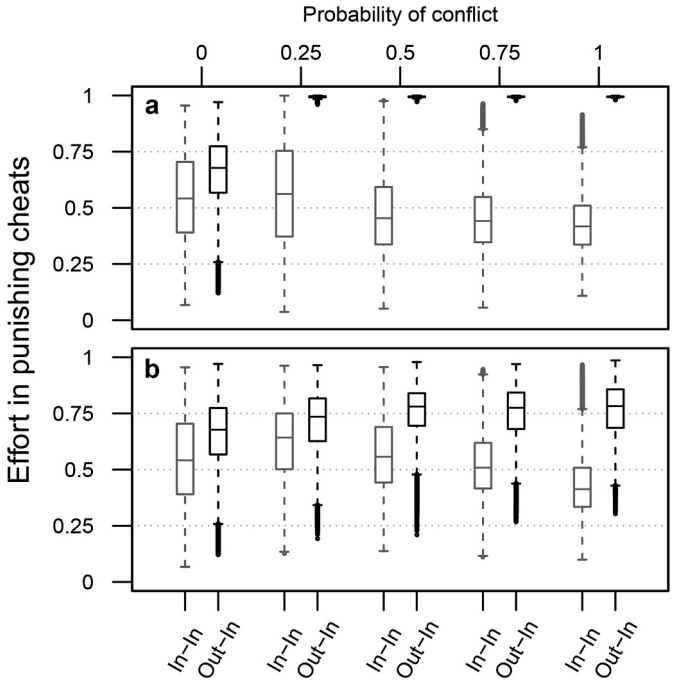

Fig. 3. Higher chances of conflict lead to stronger parochial punishment.
Long-term range of the effort in punishing cheats when conflict occurs with the neighbour group (**a**), or a random group (**b**). Data shown are from a parameter combination representative of most of the parameter range (see Fig. S14). In all panels, boxplots show the interquartile range and median over the last 500,000 generations of a simulation run. Parameters: $n = 20$, $n_g = 100$, $h = 4$, $R_{In\text{-}In} = 200$, $R_{Out\text{-}In} = 200$, $R_{In\text{-}Out} = 200$, $c = 1$, $b = 2$, $\alpha = 2$, $\beta = 6$, $\mu = 0.01$.

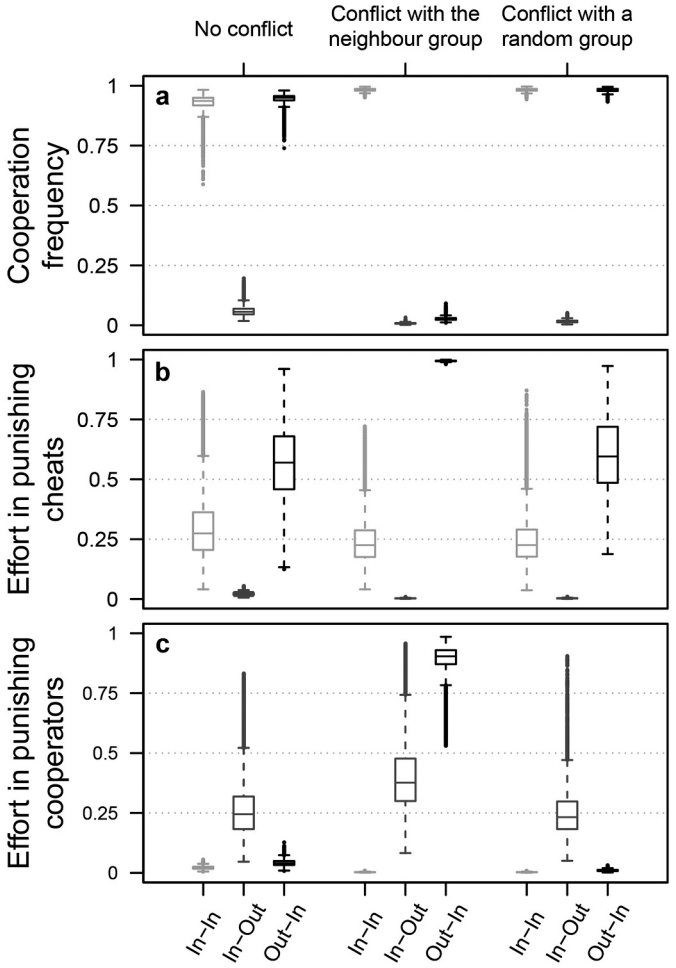

Fig. 4. Parochial punishment can evolve without conflict.
Long-term range of the frequency of cooperation (**a**), effort in punishing cheats (**b**) and cooperators (**c**) for each subgroup composition in each ecological scenario, when only the punished individuals can adjust their behaviour after punishment (instead of the whole group). In all panels, boxplots show the interquartile range and median over the last 500,000 generations of a simulation run (see Fig. S15 for a broader range of parameter values). Parameters: $n = 20$, $n_g = 100$, $h = 4$, $R_{In\text{-}In} = 200$, $R_{Out\text{-}In} = 200$, $R_{In\text{-}Out} = 200$, $c = 1$, $b = 2$, $\alpha = 2$, $\beta = 6$, $\mu = 0.01$.

cooperation and punishment, because a single act of severe punishment in a group can induce everyone to cooperate in all subsequent interactions. Therefore, to test whether our results rely on this assumption, we ran additional simulations in which the response to punishment is restricted to the punished individual only, like in previous models (Boyd & Richerson, 1992 Krasnow, Delton, Cosmides, & Tooby, 2015). As before, we find that conflict with the neighbour group leads to spiteful punishment, and conflict with a random group leads to parochial punishment (Fig. 4; Fig. S15). However, we find that parochial punishment can also evolve in the absence of intergroup conflict (when groups compete against nature) if the cooperation benefit is low (Fig. 4; Fig. S15).

The reason why parochial punishment evolves in this case is that cooperation is harder to enforce, as each punishment can convert only one cheat at a time (as opposed to all cheats at once). Many in- and out-group cheats need to be disciplined. Punishing in-group cheats is more costly to the punisher's group success ($\alpha + \beta$) than punishing out-groups (only α). These increased costs from disciplining in-group cheats select for less severe in-group punishment. Accordingly, when cooperation is easier to evolve under high benefits (b), full within-group cooperation evolves, mutant cheats are much rarer, which relaxes the selection pressure for low in-group punishment (Fig. S15). These findings suggest that when cooperation is difficult to enforce and that many punishment acts need to be done, there is selection for mild in-group punishment, leading to a group bias in punishment decisions. Furthermore, these results demonstrate that parochial punishment can evolve even in the absence of intergroup conflict.

4. Discussion

We asked which ecological conditions are likely to favour the evolution of parochial third-party punishment – the tendency to punish out-group cheats more severely than in-groups. Previous studies suggested that a parochial bias such as parochial cooperation, evolved because social groups engaged in physical conflict with each other. We found that (i) in the absence of intergroup conflict, cooperation took place both within- and between-groups, and parochial third-party punishment could evolve when cooperation was difficult to enforce and required many punitive acts, (ii) if conflict was present and occurred between groups that previously interacted with each other, within-group interactions were highly cooperative, but between-group interactions were spiteful (i.e., no cooperation and full punishment of both cheats and cooperators), suggesting that conflict with an interacting group is not sufficient for a parochial bias in punishment decisions to evolve, (iii) if groups could develop collaborative relationships with each other (i.e., high number of between-group interactions), and were more likely to come into conflict with a group other than the one they previously interacted with, then parochial punishment could evolve, between-group cooperation was high, and out-group cooperators were not punished. This behavioural pattern was the most similar to the patterns usually observed in laboratory experiments.

What are the ultimate reasons why a group bias in punishment decisions evolved in our simulation? We found that two mutually non-exclusive factors could generate this bias. A first factor was selection for

low in-group punishment. This occurred in the absence of conflict, when cooperation was difficult to enforce (when we assumed a psychology in which only the punished agent could react to punishment, as opposed to everyone in the group), so that punishing in-groups greatly undermined group success. Therefore, we predict that people will lower their punishment severity towards in-groups when many cheats need to be punished and group success is at stake. Selection for low in-group punishment also occurred in our simulation when groups engaged in physical conflict with a random group and the punishment ratio was high, so that the punishment of each rare in-group cheat could make a group lose conflict. Accordingly, we observed a decrease in the punishment severity towards in-group cheats when the probability of conflict increased. A recent experiment actually showed that people punished out-group cheats more harshly after groups either cooperated or competed with each other, and that competition increased this bias (Guo et al., 2020). While these results confirm that group competition is not a necessary condition for parochial punishment, this study also found that competition did not decrease the punishment severity towards in-group cheats. Competition even seemed to increase the harshness of in-group punishment in another experimental study (Sääksvuori, Mappes, & Puurtinen, 2011). However, a crucial difference with our setting is that competition between groups occurred before the punishment stage. So punishment had no influence on group success, which might explain why in-groups were not less severely punished when competition occurred in these studies.

A second factor that generated a group bias in punishment decisions in our simulation was selection for high punishment severity towards out-groups. Regardless of the psychology we assumed for our agents, selection for high out-group punishment occurred when groups engaged in conflict. This result is in line with empirical studies showing that direct competition between groups increased the punishment severity towards out-groups (Götte et al., 2012 Guo et al., 2020). However, when conflict occurred with a group other than the neighbour group in our simulation, a high punishment severity towards rare out-group cheats ensured that these cheats would be turned into co-operators in subsequent interactions. Being severe with out-groups also enabled individuals to weaken the neighbour group, in the rare case that the neighbour group became an enemy during conflict. These predictions remain to be empirically tested. We believe that future research will benefit from investigating different settings in which interactions within- and between-groups are repeated, and groups do not necessarily compete with the group they previously interacted with.

Our results suggest that both collaborative interactions and inter-group conflict are among the potential factors that could explain the evolution of parochial punishment in early humans. Both conditions were probably common during our evolutionary history (Lahr et al., 2016 Pisor & Surbeck, 2019). Thus, our results support the plausibility of the hypothesis that selection could have favoured a psychology which takes into account group affiliation when making cooperative and punitive decisions. Such evolved psychology could explain the behaviours observed both in laboratory experiments and in real life when people face group settings. However, there are other potential mechanisms which could favour the evolution of parochial punishment. An example is how different reputation dynamics within- and between-groups could potentially lead to an increased severity in the punishment of out-group cheats. Such mechanisms have been shown to affect parochial cooperation in the absence of punishment (Masuda, 2012 Matsuo, Jusup, & Iwasa, 2014 Nakai, 2014 Nakamura & Masuda, 2012). Hence, combining individual reputations with punishment both within- and between-groups could lead to more insight into the evolution parochial punishment.

Our findings also suggest that between-group cooperation could be increased in real-life by making the potential to form cooperative relationships with out-group members more visible, and by reducing the perception that competition occurs with interacting out-groups. These manipulations could help reduce intergroup biases, which often lead to negative consequences in human societies. In accordance with these predictions, an experimental study showed that out-group bias can be reduced when members of different social groups have to collaborate to achieve a common goal (Adachi, Hodson, Willoughby, & Zanette, 2015). Furthermore, we also predict that behaviour towards out-groups can be shaped by how in- and out-groups react to it. In our simulation, when only the donor was an out-group, cooperation levels were as high as when all three players came from the same group. Therefore, instead of cooperating less with out-groups in general, we expect people to take the third-party's group affiliation into account as well. In line with this prediction are the results from an experimental study (Bernhard, Fischbacher, & Fehr, 2006). In a similar game as in our simulation, they found that most donors (dictators in their case) expected maximal punishment after cheating (transfers of 0) if the punisher was from the same group as that of the recipient. With such expectations, it is likely that donors would cooperate fully.

Finally, three assumptions in our model could be amended in future research. First, we assumed that agents could only react to the last punishment received, and could not, for example, keep track of the average punishment received by either themselves or their group members. As we have shown, the type of punishment responsiveness modelled is likely to affect whether a parochial bias in punishment emerges. Hence, a future research direction would be to investigate when parochial punishment evolves under such more sophisticated psychologies. Second, we assumed that conflicts occur exogenously (defined by a fixed probability), rather than endogenously as the result of the individuals' decisions. However, conflicts in nature are likely to be triggered when there are tensions or a lack of cooperation between interacting groups. A group's tendency to engage in conflict could therefore be made contingent on the amount of between-group cheating and/or antisocial punishment to better reflect natural conditions. Third, we assumed fixed, rather than flexible group membership over an individual's lifetime. It has been argued that real human groups can often be unstable (Palla, Barabási, & Vicsek, 2007 Rand et al., 2009), even in small-scale societies (Marlowe, 2005). Two theoretical studies investigated how parochial cooperation could evolve without conflict when groups are ephemeral, but punishment was not included (Fu et al., 2012 Gross & De Dreu, 2019). Therefore, a potential extension of our model would be to test how dynamic group compositions shape the evolution of parochial punishment.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

We thank Coren Apicella, Simon Powers and our anonymous reviewers for comments. This work was supported by grants from the Typhaine Foundation (CH) to Daria Knoch. Simulations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2020.10.002.

References

Aaldering, H., Ten Velden, F. S., van Kleef, G. A., & De Dreu, C. K. W. (2018). Parochial cooperation in nested intergroup dilemmas is reduced when it harms out-groups. *Journal of Personality and Social Psychology, 114*(6), 909–923. https://doi.org/10.1037/pspi0000125.

Adachi, P. J. C., Hodson, G., Willoughby, T., & Zanette, S. (2015). Brothers and sisters in arms: Intergroup cooperation in a violent shooter game can reduce intergroup bias. *Psychology of Violence, 5*, 455–462. https://doi.org/10.1037/a0037407.

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140*(6), 1556–1581. https://doi.org/10.1037/a0037737.

Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping, 33*, 1452–1469. https://doi.org/10.1002/hbm.21298.

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R. R., & Knoch, D. (2014). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Social Cognitive and Affective Neuroscience, 9*(5), 653–660. https://doi.org/10.1093/scan/nst023.

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review, 96*(2), 217–221. https://doi.org/10.1257/000282806777212594.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442*, 912–915. https://doi.org/10.1038/nature04981.

Böhm, R., Rusch, H., & Baron, J. (2018). The psychology of intergroup conflict: A review of theories and measures. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2018.01.020.

Bowles, S. (2008). Conflict: altruism's midwife. *Nature, 456*, 326–327.

Bowles, S. (2009). Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science, 324*, 1293–1298. https://doi.org/10.1126/science.1168112.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*(3), 171–195. https://doi.org/10.1016/0162-3095(92)90032-Y.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*, 429–444. https://doi.org/10.1111/0022-4537.00126.

Choi, J. K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science, 318*, 636–640. https://doi.org/10.1126/science.1144237.

De Dreu, C. K. W., Balliet, D., & Halevy, N. (2014). Parochial cooperation in humans: Forms and functions of self-sacrifice in intergroup conflict. In A. Elliot (Vol. Ed.), *Advances in motivation science. Vol. 1. Advances in motivation science* (pp. 1–47). New York: Elsevier. https://doi.org/10.1016/bs.adms.2014.08.001.

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., ... Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*, 1408–1411. https://doi.org/10.1126/science.1189047.

Delton, A. W., & Krasnow, M. M. (2017). The psychology of deterrence explains why group membership matters for third-party punishment. *Evolution and Human Behavior, 38*, 734–743. https://doi.org/10.1016/j.evolhumbehav.2017.07.003.

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science, 321*, 1844–1849. https://doi.org/10.1126/science.1155805.

Everett, J. A. C., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience, 9*, 15. https://doi.org/10.3389/fnbeh.2015.00015.

Fu, F., Tarnita, C. E., Christakis, N. A., Wang, L., Rand, D. G., & Nowak, M. A. (2012). Evolution of in-group favoritism. *Scientific Reports, 2*, 460. https://doi.org/10.1038/srep00460.

Gao, S., Wu, T., Nie, S., & Wang, L. (2015). Emergence of parochial altruism in well-mixed populations of multiple groups. *Physics Letters A, 379*, 2311–2318. https://doi.org/10.1016/j.physleta.2015.07.032.

Gao, S., Wu, T., & Wang, L. (2015). Emergence of parochial altruism in well-mixed populations. *Physics Letters A, 379*, 333–341. https://doi.org/10.1016/j.physleta.2014.11.035.

García, J., & van den Bergh, J. C. J. M. (2011). Evolution of parochial altruism by multilevel selection. *Evolution and Human Behavior, 32*, 277–287. https://doi.org/10.1016/j.evolhumbehav.2010.07.007.

Götte, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement. *American Economic Review, 96*, 212–216. https://doi.org/10.1257/000282806777211658.

Götte, L., Huffman, D., Meier, S., & Sutter, M. (2012). Competition between organizational groups: Its impact on altruistic and antisocial motivations. *Management Science, 58*(5), 948–960. https://doi.org/10.1287/mnsc.1110.1466.

Gross, J., & De Dreu, C. K. W. (2019). The rise and fall of cooperation through reputation and group polarization. *Nature Communications, 10*, 776. https://doi.org/10.1038/s41467-019-08727-8.

Guo, R., Ding, J., & Wu, Z. (2020). How intergroup relation moderates group bias in third-party punishment. *Acta Psychologica, 205*, 103055. https://doi.org/10.1016/j.actpsy.2020.103055.

Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making, 25*, 188–195. https://doi.org/10.1002/bdm.726.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*, 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109.

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 12710–12715. https://doi.org/10.1073/pnas.1402280111.

Jordan, J. J., & Rand, D. G. (2017). Third-party punishment as a costly signal of high continuation probabilities in repeated games. *Journal of Theoretical Biology, 421*, 189–202. https://doi.org/10.1016/j.jtbi.2017.04.004.

Konrad, K. A., & Morath, F. (2012). Evolutionarily stable in-group favoritism and out-group spite in intergroup conflict. *Journal of Theoretical Biology, 306*, 61–67. https://doi.org/10.1016/j.jtbi.2012.04.013.

Koopmans, R., & Rebers, S. (2009). Collective action in culturally similar and dissimilar groups: An experiment on parochialism, conditional cooperation, and their linkages. *Evolution and Human Behavior, 30*, 201–211. https://doi.org/10.1016/j.evolhumbehav.2009.01.003.

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2015). Group cooperation without group selection: Modest punishment can recruit much cooperation. *PLoS One, 10*. https://doi.org/10.1371/journal.pone.0124561 e0124561.

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science, 27*, 405–418. https://doi.org/10.1177/0956797615624469.

Krasnow, M. M., Delton, A. W., Tooby, J., & Cosmides, L. (2013). Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation. *Scientific Reports, 3*, 1747. https://doi.org/10.1038/srep01747.

Lahr, M. M., Rivera, F., Power, R. K., Mounier, A., Copsey, B., Crivellaro, F., ... Foley, R. A. (2016). Inter-group violence among early Holocene hunter-gatherers of West Turkana, Kenya. *Nature, 529*, 394–398. https://doi.org/10.1038/nature16477.

Lehmann, L. (2011). The demographic benefits of belligerence and bravery: Defeated group repopulation or victorious group size expansion? *PLoS One, 6*, Article e21437. https://doi.org/10.1371/journal.pone.0021437.

Lehmann, L., & Feldman, M. W. (2008). War and the evolution of belligerence and bravery. *Proceedings of the Royal Society B: Biological Sciences, 275*, 2877–2885. https://doi.org/10.1098/rspb.2008.0842.

Liu, Y., Bian, X., Hu, Y., Chen, Y.-T., Li, X., & Di Fabrizio, B. (2018). Intergroup bias influences third-party punishment and compensation: In-group relationships attenuate altruistic punishment. *Social Behavior and Personality: An International Journal, 46*, 1397–1408. https://doi.org/10.2224/sbp.7193.

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology, 14*, 54–67. https://doi.org/10.1002/evan.20046.

Martin, J., Young, L., & McAuliffe, K. (2020). The impact of group membership on punishment versus partner choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/5qr32.

Masuda, N. (2012). Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation. *Journal of Theoretical Biology, 311*, 8–18. https://doi.org/10.1016/j.jtbi.2012.07.002.

Masuda, N., & Fu, F. (2015). Evolutionary models of in-group favoritism. *F1000Prime Reports, 7*. https://doi.org/10.12703/p7-27.

Matsuo, T., Jusup, M., & Iwasa, Y. (2014). The conflict of social norms may cause the collapse of cooperation: Indirect reciprocity with opposing attitudes towards in-group favoritism. *Journal of Theoretical Biology, 346*, 34–46. https://doi.org/10.1016/j.jtbi.2013.12.018.

McAuliffe, K., & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 371*, 20150073. https://doi.org/10.1098/rstb.2015.0073.

McDonald, M. M., Navarrete, C. D., & van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 367*, 670–679. https://doi.org/10.1098/rstb.2011.0301.

Mussweiler, T., & Ockenfels, A. (2013). Similarity increases altruistic punishment in humans. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 19318–19323. https://doi.org/10.1073/pnas.1215443110.

Nakai, Y. (2014). In-group favoritism due to friend selection strategies based on fixed tag and within-group reputation. *Rationality and Society, 26*, 320–354. https://doi.org/10.1177/1043463114523714.

Nakamura, M., & Masuda, N. (2012). Groupwise information sharing promotes ingroup favoritism in indirect reciprocity. *BMC Evolutionary Biology, 12*, 213. https://doi.org/10.1186/1471-2148-12-213.

Ockenfels, A., & Werner, P. (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior and Organization, 108*, 453–462. https://doi.org/10.1016/j.jebo.2013.12.003.

Okasha, S. (2007). *Evolution and the Levels of Selection. Evolution and the Levels of Selection.* https://doi.org/10.1093/acprof:oso/9780199267972.001.0001.

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin, 112*, 536–552. https://doi.org/10.1037/0033-2909.112.3.536.

Palla, G., Barabási, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*, 664–667. https://doi.org/10.1038/nature05670.

Pisor, A. C., & Surbeck, M. (2019). The evolution of intergroup tolerance in non-human primates and humans. *Evolutionary Anthropology, 28*, 210–223.

Puurtinen, M., Heap, S., & Mappes, T. (2015). The joint emergence of group competition and within-group cooperation. *Evolution and Human Behavior, 36*, 211–217. https://doi.org/10.1016/j.evolhumbehav.2014.11.005.

Rand, D. G., Pfeiffer, T., Dreber, A., Sheketoff, R. W., Wernerfelt, N. C., & Benkler, Y. (2009). Dynamic remodeling of in-group bias during the 2008 residential election. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 6187–6191. https://doi.org/10.1073/pnas.0811552106.

Romano, A., Balliet, D., Yamagishi, T., & Liu, J. H. (2017). Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences, 114*, 12702–12707. https://doi.org/10.1073/pnas.1712921114.

Roos, P., Gelfand, M., Nau, D., & Carr, R. (2013). High strength-of-ties and low mobility enable the evolution of third-party punishment. *Proceedings of the Royal Society B: Biological Sciences, 281*, 20132661. https://doi.org/10.1098/rspb.2013.2661.

Rusch, H. (2014). The evolutionary interplay of intergroup conflict and altruism in humans: A review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society B: Biological Sciences, 281*, 20141539. https://doi.org/10.1098/rspb.2014.1539.

Sääksvuori, L., Mappes, T., & Puurtinen, M. (2011). Costly punishment prevails in intergroup conflict. *Proceedings of the Royal Society B: Biological Sciences, 278*(1723), 3428–3436. https://doi.org/10.1098/rspb.2011.0252.

Schaub, M. (2017). Threat and parochialism in intergroup relations: Lab-in-the-field evidence from rural Georgia. *Proceedings of the Royal Society B: Biological Sciences, 284*, 20171560. https://doi.org/10.1098/rspb.2017.1560.

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior, 35*, 169–175. https://doi.org/10.1016/j.evolhumbehav.2013.12.006.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149–178. https://doi.org/10.1002/ejsp.2420010202.

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences, 17*, 585–607. https://doi.org/10.1017/S0140525X00036104.

Yudkin, D. A., Rothmund, T., Twardawski, M., Thalla, N., & Van Bavel, J. J. (2016). Reflexive intergroup bias in third-party punishment. *Journal of Experimental Psychology: General, 145*, 1448. https://doi.org/10.1037/xge0000190.